Author Reply to Reviewer Comments

RC1 Reply

Dear Referee,

We thank you for the thorough, detailed, and constructive review. Your comments have helped us substantially improve the clarity and scientific framing of the manuscript. We address each comment below in the order presented.

General comment

Reviewer comment

The manuscript reports on the feasibility of predicting INP concentrations measured at −31°C and −32°C during the HyICE-2018 campaign in the boreal forest from 84 complementary variables measured at the site. The paper's honest takeaway is that no variable strongly predicts INP in the boreal winter, and only moderate skill ($R^2 \approx 0.5$) is achieved in spring/summer. This is an important negative result and is nicely summarized at the end of the introduction, telling the community that even many co-located variables at one of the best instrumented stations cannot explain INP variability in general. The authors could commit more fully to this message. The paper is framed around "predicting INP using machine learning" yet the conclusion effectively states that strong links remain "abstruse." This is a cautionary study about the limits of data-driven approaches for INP prediction, which is arguably more useful to the community than a modest positive result would be. However, the paper presents itself as a machine learning study, but the ML algorithms are used exclusively as importance ranking tools, not for predictive modeling. The more closely investigated relations (Table 1) are simple power-law fits with only two parameters, no different from classical empirical fitting. There is no actual ML-based predictive model evaluated with train/test splits. With data from 84 variables, a proper ML regression model (e.g., gradient boosting or random forest regressor) could be benchmarked against the power-law fits. As the paper stands, the title overstates the ML contribution.

Response

We thank the referee for this important observation. We fully agree that the ML algorithms in this study were applied as importance-ranking and pattern-exploration tools rather than to build an evaluated predictive model in the conventional sense (i.e., with train/test data splits and cross-validated performance metrics). The title "Predicting…" does not accurately reflect this scope.

Accordingly, we have changed "Predicting" to "Exploring" in the title. We have also added a clarifying sentence in the introduction (Sect. 2.2) that explicitly states the ML methods were used for feature importance ranking and hypothesis generation, not for constructing a validated predictive model. We agree with the referee that the primary scientific message—that even with more than 500 co-located variables at one of the most heavily instrumented stations globally, strong links to INP concentrations remain elusive—is a valuable cautionary result, and we have committed more fully to this framing in the revised abstract, introduction, and conclusions.

It is important to note that although an original goal was to develop a trained predictive model, based on our explorations, physical intuition and previous results we did not construct a new ML regression model with train/test evaluation because it could not be justified based on the strength of connections observed.

Change in manuscript

- Title: changed *"Predicting"* to *"Exploring."* Following further input from Reviewer 2, the final title is: *"Exploring Ice Nucleation Particle concentrations in a Boreal Environment: limits of machine-learning-assisted variable screening."* This reflects input from both reviewers: "Exploring" addresses the framing concern raised here, while "concentrations" (replacing "properties") and the "limits of..." subtitle address Reviewer 2's observation that only one INP property is examined and that the null result deserves prominence.
- Added a clarifying sentence to Sect. 2.2: *"It is important to note that these machine learning methods were applied here as importance-ranking and exploratory tools, not to build or evaluate a predictive model; no train/test data splits were applied."*
- Abstract and conclusion reframed to foreground the cautionary message (see specific edits below).

Specific comments

Line 15 (abstract)

Reviewer comment

Specify how the results underscore the need for site-specific parameterizations and suggest on which variables parameterizations could instead be based, considering that none of the 84 variables included here work.

Response

We agree that the abstract conclusion sentence should be more specific. Although no single variable provides strong predictive power across the full dataset, the spring/summer PINCii data show that local biogenic and chemical proxies (fluorescent particle concentrations and nitrate aerosol mass) yield adjusted $R^2 \approx 0.5$. We have added a mention of these proxies to the final abstract sentence.

Change in manuscript

Changed the final abstract sentence from *"These results underscore the need for site-specific parameterizations to capture INP variability in the complex boreal environments."* to *"These results underscore the need for site-specific parameterizations to capture INP variability in the complex boreal environments; local biogenic and chemical proxies, such as fluorescent particle concentrations and nitrate aerosol mass, emerge as the most promising predictors for the spring and summer period."*

Line 20

Reviewer comment

Explain why the coexistence is inherently unstable. In the same sentence it is mentioned that mixed-phase clouds persist for days, which seems to contradict this statement.

Response

"Inherently unstable" refers to thermodynamic instability: below 0°C, ice is the thermodynamically stable phase and supercooled liquid water is in a metastable state. The apparent contradiction with long-lived mixed-phase clouds is resolved by the distinction between thermodynamic instability and kinetic stability. Spontaneous ice nucleation faces a high kinetic energy barrier (homogeneous nucleation requires temperatures below approximately −38°C), so supercooled liquid water persists until an ice-nucleating particle (INP) catalyses freezing by lowering this barrier. Mixed-phase clouds can therefore persist for hours to days despite being thermodynamically unstable, precisely because efficient INPs are rare. We have added a brief parenthetical to make this distinction explicit.

Change in manuscript

Changed *"unstable co-existence of ice and liquid water"* to *"thermodynamically metastable co-existence of ice and supercooled liquid water (below 0°C, ice is the stable phase but a high kinetic barrier to nucleation sustains liquid water in the absence of efficient INPs)"*.

Line 22

Reviewer comment

The Arctic is considered a high-latitude region, not a region beyond high latitudes.

Response

We thank the referee for this correction. "Beyond high-latitudes" is geographically imprecise. We have corrected the phrasing.

Change in manuscript

Changed *"Beyond high-latitudes"* to *"At high latitudes, particularly in the Arctic"*, mixed-phase clouds play out-sized roles in regulating climate…

Line 23

Reviewer comment

Summarize what the underlying amplifying feedbacks are.

Response

We have added brief examples of the relevant Arctic feedbacks.

Change in manuscript

Changed *"especially in the Arctic where underlying feedbacks have amplifying effects"* to *"where amplifying feedbacks—such as the ice-albedo feedback and the water vapour–temperature amplification characteristic of polar warming—have particularly pronounced effects"*.

Line 26

Reviewer comment

Summarize how CCN and INP are fundamental and to which cloud processes.

Response

We have specified the relevant cloud processes.

Change in manuscript

Changed *"cloud processes"* to *"cloud formation, precipitation efficiency, and radiative properties"*.

Line 27

Reviewer comment

Clarify the difference between INP occurrence and abundance.

Response

We have added a brief parenthetical to define both terms.

Change in manuscript

Changed *"occurrence and abundance"* to *"occurrence (whether INPs are present at all) and abundance (how many are present per unit volume of air)"*.

Line 33

Reviewer comment

Providing some more details on the instrumentation, specifically about the instruments measuring the 84 variables used in this study, would be helpful. This could be done in a supplementary table including the instrument name, brand, sampling frequency, and volume/flow.

Response

Comprehensive instrument metadata for all SMEAR II variables is documented in the SmartSMEAR database (Junninen et al., 2009) and the station overview (Hari and Kulmala, 2005). For the HyICE-2018 campaign instruments specifically, a detailed instrument table including names, design specifications, and operating conditions is provided in Brasseur et al. (2022, their Table 1 and Appendix). We have added a sentence in Sect. 2.2 directing readers to these resources.

A dedicated supplementary table is not included because complete metadata are accessible via the SmartSMEAR online portal, making a separate table redundant.

Change in manuscript

Added the following sentence to Sect. 2.2: *"Detailed instrument specifications for the HyICE-2018 campaign instruments are provided in Brasseur et al. (2022); metadata for the full SMEAR II monitoring suite are available via the SmartSMEAR portal (Junninen et al., 2009)."*

Lines 50 and 75

Reviewer comment

Specify the time resolution with which the CFDCs measure INP concentrations and explain why the data are averaged over 20 min or 1 hour (Figure 3). Figure 1 shows INP concentrations often reaching hundreds per liter. With a sample flow of 1 L min^{-1} there should be plenty of signal in 1 min averages or even 10 s averages, which would substantially increase the number of INP data points available for analysis. Additionally, correlations of ambient measurements depend on the time resolution or averaging interval. The temporal scale at which a correlation is seen also identifies the scale of the process that drives the changes in variables. Investigating correlations at different time resolutions, which seems possible with this dataset, could be interesting for the very variable INP concentrations. Such an analysis could be added, and the dependence of correlation analyses on the temporal data resolution should be discussed.

Response

Both PINC and PINCii share an identical measurement cycle: a 5-minute background (particle-free) period followed by a 15-minute ambient sampling period, cycling continuously (i.e., . . . 5 min background − 15 min ambient − 5 min background − 15 min ambient . . .). Each INP concentration value is derived by subtracting the interpolated background count from the ambient-period count during the 15-minute window, yielding one data point every 20 minutes. The 15-minute ambient window is therefore the sampling window; the 20-minute figure is the data cadence, not the sampling window. This background–ambient cycle is the instruments' fixed operational protocol, not a post-hoc averaging choice. At a sample flow rate of 1 L min^{-1}, the 15-minute window provides a 15 L sample volume per data point. During the PINC winter deployment, where INP concentrations at $T_l = -31$°C were typically <1 INP L^{-1}, this volume ensures that most sampling intervals yield at least a few particle counts sufficient for a reliable concentration estimate.

The referee notes that Fig. 1 shows periods—primarily during PINCii spring/summer operation—where INP concentrations reach hundreds per liter, for which even a 1-minute ambient window would yield ample particle counts. However, the background–ambient cycle is the fixed operational protocol of both instruments and cannot be shortened without fundamentally redesigning the background subtraction scheme. The cycle was held constant throughout the campaign to ensure consistency across all data points.

The hourly mean data used for the Pearson correlation analysis in Fig. 3 were chosen to align with the coarser time resolution of some complementary datasets. We have explicitly stated this reasoning in the Fig. 3 caption and confirmed that repeating the analysis with the native 20-minute data cadence yields qualitatively similar Pearson correlations.

We acknowledge the referee's conceptual point that the temporal scale of a detected correlation can reveal the timescale of the driving process. The qualitatively similar correlations at 20-minute and hourly resolution suggest that the dominant co-varying processes operate on timescales longer than 20 minutes (diurnal or synoptic scale), consistent with the broad seasonal structure visible in Fig. 1. A systematic multi-scale temporal correlation analysis at all available resolutions would, however, constitute a substantial undertaking; we have added a brief discussion of these implications.

Change in manuscript

- Added to Sect. 2.3: *"Both PINC and PINCii operate with a 15-minute ambient sampling window bracketed by 5-minute background (particle-free) periods, yielding one INP data point every 20 minutes. At a sample flow rate of 1 L min^{-1}, the 15 L sample volume per data point is sufficient for statistically meaningful INP counts even at the low concentrations observed during winter (<1 L^{-1})."*
- Updated time-series description (Sect. 3.1) to: *"...each with a 15-minute ambient sampling window (one data point every 20 minutes) ..."* and *"Each point represents a 15-minute ambient sampling period (with 5-minute background periods before and after, yielding one data point every 20 minutes)."*
- Added to the Fig. 3 discussion and caption: reasoning for the use of hourly means (alignment with coarser complementary datasets), confirmation that the native 20-minute data cadence yields qualitatively similar Pearson correlations, and a note on the implications for campaign-based INP correlations reported in the literature.

Section 2.2

Reviewer comment

Currently, the ML techniques are not explained in this section. Consider changing the section title to "Complementary Data Selection." Add more details about how datasets were processed for the analysis. Add a separate methodology section about ML algorithms, feature selection criteria, to make the approach reproducible.

Response

We agree that the ML methodology deserves more explicit documentation. We have added a paragraph to Sect. 2.2 that describes: (i) the algorithms applied (random forest and decision trees for feature importance ranking; principal component analysis and K-means clustering for dimensionality reduction and pattern exploration); (ii) that no train/test data split was applied, as the purpose was exploratory importance ranking; and (iii) how datasets were harmonised to a common time base (20-minute resolution, with coarser-resolution variables assigned to the nearest 20-minute interval).

We elect to update the section title "Complementary Data and Machine Learning" and maintain a single section that covers both the data and the ML framework.

Change in manuscript

Added the following paragraph to Sect. 2.2:

> *"The machine learning analysis employed random forest and decision tree models to derive feature importance rankings, and principal component analysis (PCA) together with K-means clustering for dimensionality reduction. These algorithms were applied as exploratory tools to rank variables by their statistical association with INP concentrations; no train/test data splitting or cross-validation was performed. Prior to analysis, all variables were harmonised to a common 20-minute time base; variables measured at coarser time resolution were assigned to the nearest 20-minute interval, and those measured at finer resolution were averaged."*

Line 73

Reviewer comment

Explain which hypotheses are tested to illuminate the sources and mechanisms.

Response

We have clarified the scientific hypotheses that motivated the variable selection and analysis in Sect. 2.3.

Change in manuscript

Added the following sentence to Sect. 2.3: *"Specifically, we test whether new particle formation (NPF) events generate particles that grow into the INP-relevant size range (>0.5 μm); whether primary biological aerosol (proxied by fluorescent particle counts from the WIBS) is a dominant INP source in this environment; and whether aerosol chemical markers—including black carbon, nitrate, and organic aerosol mass—can serve as practical INP proxies."*

Line 76

Reviewer comment

Clarify what is meant by "straightforwardly intercompared."

Response

The phrase was intended to indicate that PINC and PINCii share a common operating principle (parallel-plate CFDC design) and can thus be compared on similar methodological terms. The word "straightforwardly" is unnecessary and potentially misleading. We have simplified the phrasing.

Change in manuscript

Changed *"are somewhat straightforwardly intercompared as demonstrated by"* to *"have been directly compared by"*.

Section 2.3

Reviewer comment

Clarify why the details on the CFDC chambers are relevant for this paper. None of the details are referred to later.

Response

The CFDC technical specifications (chamber dimensions, flow rates, temperature and humidity conditions) are standard documentation for peer-reviewed CFDC-based INP papers, enabling reproducibility and comparison with other CFDC studies in the literature. More importantly, the operating conditions ($T_l = -31$°C for PINC, $T_l = -32$°C for PINCii, $\mathrm{RH}_w = 105\%$) define the ice nucleation activation threshold and the temperature at which the reported INP concentrations were measured—directly relevant to interpreting all results. We have added a brief sentence at the opening of Sect. 2.3 clarifying this purpose.

Change in manuscript

Added the following sentence at the opening of Sect. 2.3: *"The following instrument specifications are provided to establish the conditions under which INP concentrations were measured (particularly the lamina temperature and humidity, which define the INP activation threshold) and to facilitate comparison with other CFDC-based INP studies."*

Line 107

Reviewer comment

Contrary to what is stated here, Brasseur et al., 2022 mention a sampling window of 15 min for PINCii.

Response

We thank the referee for this correction. Brasseur et al. (2022) specify a 15-minute ambient sampling window for PINCii. In fact, PINC follows the same 5-minute background + 15-minute ambient cycle, so both instruments have a 15-minute sampling window and produce one data point every 20 minutes. The original phrase "20 minute sampling windows" incorrectly equated the data cadence (one point per 20 min) with the sampling window (15 min ambient period). We have corrected all occurrences in the manuscript to accurately describe the 15-minute ambient sampling window and the 20-minute data cadence.

Change in manuscript

- Corrected *"with 20 minute sampling windows"* to *"each with a 15-minute ambient sampling window (one data point every 20 minutes) (Brasseur et al., 2022)"*.
- Corrected *"Each point represents a 20 minute sampling window"* to *"Each point represents a 15-minute ambient sampling period (with 5-minute background periods before and after, yielding one data point every 20 minutes)"*.

Figure 1

Reviewer comment

Indicate the temperature at which INP concentrations were measured. Check the units in panel c. In panel e, the last line is shown with very weak colours.

Response

We thank the reviewer for catching the unit error in panel (c). We have updated Figure 1 and its caption to address all three points: (i) the measurement temperatures (−31°C for PINC, −32°C for PINCii) are now stated in the panel (a) caption; (ii) the units in panel (c) have been corrected to #/cm^3; and (iii) regarding the weak colours in panel (e), the colour scheme follows the standard AMS convention (chloride: purple; ammonia: yellow; sulfate: red; nitrate: blue; organics: green), which we have now noted explicitly in the caption.

Change in manuscript

Figure 1 caption updated: (i) added *"INP concentrations were measured at $T_l = -31\,°C$ (PINC) and $T_l = -32\,°C$ (PINCii)"* to the panel (a) description; (ii) added *"(#/cm^3)"* as the unit in the panel (c) description; and (iii) added *"(colours follow AMS convention)"* to the panel (e) description.

Lines 119–120

Reviewer comment

Clarify why NPF is relevant here.

Response

NPF events generate large numbers of freshly nucleated particles that can grow into the size range relevant for ice nucleation during sustained growth events lasting several hours. Additionally, NPF events are a well-known feature of the SMEAR II station and could influence the total particle loading and composition in ways that are relevant to the INP analysis. We have added a brief explanatory sentence.

Change in manuscript

Added the following sentence to Sect. 3.1: *"New particle formation (NPF) events are notable here because freshly nucleated particles can grow into the INP-relevant size range during sustained growth episodes, potentially contributing to the total INP-active aerosol population."*

Line 122

Reviewer comment

What relationships can be observed in Figure 1 that motivate the analysis?

Response

We have revised the introductory sentence to Sect. 3.2 to specify the visual patterns that motivate the more systematic analysis: the seasonal progression in temperature and snow depth, the spring/summer increase in fluorescent biological particles, the shift in aerosol chemical composition (increasing organic fraction), and the apparent co-variability of INP with some of these tracers.

Change in manuscript

Changed the opening of Sect. 3.2 from *"The relationships suggested by the time series in Fig. 1 motivate a more objective evaluation. . . "* to *"The seasonal progression visible in Fig. 1—the transition in temperature and snow depth, the spring/summer increase in fluorescent biological particles, the shift in aerosol chemical composition (increasing organic fraction, panel e), and the apparent co-variability of INP with these tracers—motivates a more objective evaluation. . . "*

Lines 132–134

Reviewer comment

Repetition from Sect. 2.2.

Response

We thank the referee for identifying this duplication. The description of the dimensional reduction criteria (exclusion of NaN-heavy, low-variability, and redundant variables) was already presented in Sect. 2.2 and has been removed from Sect. 3.2.

Change in manuscript

Removed the following paragraph from Sect. 3.2 (it duplicated material already presented in Sect. 2.2):

> *"Variables were excluded from the original set of 509 if they contained excessive numbers of NaN values, exhibited very low variability (i.e., were nearly constant), or were effectively redundant (for example, the same parameter, such as temperature, measured at different heights without meaningful differences). Particle size distribution measurements were consolidated into number concentrations over selected size ranges. In addition, several features were found to be strongly correlated (based on Pearson correlation), indicating redundant information; one example is the close correspondence between highly oxygenated organic molecule (HOM) monomers and organic nitrate."*

Line 139

Reviewer comment

Use "INP concentration" instead of "ice nucleation activity."

Response

We agree and have replaced both occurrences in Sect. 2.2.

Change in manuscript

- Changed *"Although ice nucleation activity"* to *"Although INP concentration"* (first occurrence in Sect. 2.2).
- Changed *"that can be used to follow ice nucleation activity"* to *"that can be used to follow INP concentration"* (second occurrence in Sect. 2.2).

Line 140

Reviewer comment

The references attribute bio-INP to much lower concentrations at higher temperatures than below −30°C. Provide a supporting citation for biological particles contributing substantially at low temperatures.

Response

The referee correctly notes that primary biological INPs (fungal spores, pollen, bacteria) are typically most active at temperatures above approximately −20°C, and contributions at −31/−32°C are less well established. We have revised the statement to clarify that fluorescent particles serve as a tracer for biological aerosol that may include INP-active species—rather than claiming they are the dominant INP at our measurement temperatures—consistent with the observational approach in Paramonov et al. (2020) and Schneider et al. (2021) at the same site.

Change in manuscript

Changed *"biological particles that yield a fluorescence signal are known to be some of the most abundant INPs in many settings"* to *"biological particles that yield a fluorescence signal serve as a tracer for primary biological aerosol, which may include INP-active species, and are among the most studied INP types in forest environments (Murray et al., 2012; Morris et al., 2014; Proske et al., 2025)"*.

Line 144

Reviewer comment

Only half of the variables are listed in Figure 2, while others are selected by intuition. Specify for each variable based on which information or hypothesis it was selected.

Response

We have added explicit justification for each of the six variables shown in Fig. 3. Fluorescent particle concentrations and >0.5 µm number concentration are established predictors used in the DeMott (2010) and Tobo (2013) parameterizations; particle mass is a proxy for total aerosol loading; black carbon emerged as a top-ranked variable in the random forest analysis; nitrate ranked highly and co-varies with biogenic tracers; organic aerosol mass is motivated by Hyytiälä's well-documented biogenic aerosol seasonality (Schneider et al., 2021).

Change in manuscript

Added the following sentence to Sect. 3.2: *"The six variables shown in Fig. 3 were selected on two complementary grounds: (i) high importance rank in the random forest analysis (fluorescent particles, black carbon, nitrate, particle mass); and (ii) established physical or empirical connection to INP in prior literature or at this specific site (>0.5 µm concentration per DeMott et al. (2010) and Tobo et al. (2013); organic aerosol mass per the biogenic seasonality documented by Schneider et al. (2021))."*

Lines 148ff

Reviewer comment

That BC ranks highly in Fig. 2 and yields the highest adjusted R^2 of all predictors for PINCii (Table 1), yet is a poor INP in laboratory studies, is a provocative finding that could be explored beyond noting that it is surprising and pointing to aging and oxidation as possible explanations. It could be mentioned that Paramonov et al. (2020) also reported that BC correlated well with INP concentrations on a short timescale (their Fig. 5).

Response

We thank the referee for pointing to this reference. We have added a citation to Paramonov et al. (2020) and a sentence acknowledging that their short-timescale BC–INP correlation at the same site independently supports our finding, and expanded the discussion of possible mechanisms (aging, oxidation, and mixing with organic material).

Change in manuscript

Added the following sentences to Sect. 3.2: *"Notably, Paramonov et al. (2020) also reported a positive correlation between BC and INP concentrations at short timescales during the HyICE-2018 campaign at the same site (their Fig. 5), independently supporting this finding. Possible mechanisms include the enhancement of BC's ice-nucleating ability through atmospheric aging, oxidation, and coating with organic material (DeMott et al., 1990; Mahrt et al., 2020a,b)."*

General suggestion: connections to previous HyICE articles

Reviewer comment

In general, the paper would be strengthened if connections to findings from previous HyICE articles were integrated in more detail.

Response

We thank the referee for this general suggestion and agree that tighter integration with the other HyICE-2018 publications strengthens the manuscript. In the revised text we have (i) added a citation to and discussion of Paramonov et al. (2020) in the context of the BC–INP correlation finding (see response to lines 148ff above); (ii) retained and where appropriate expanded citations to Schneider et al. (2021), Brasseur et al. (2022), Brasseur et al. (2024), and Vogel et al. (2024) in the introduction and results sections; and (iii) added a sentence noting that the better Tobo (2013) agreement in Brasseur et al. (2022, their Fig. 8 and consolidated 28 March summary in Fig. 11) reflects a focused intercomparison subset rather than full campaign conditions (see response to Figure 5 below). We believe these additions place our results more explicitly in the context of the broader HyICE-2018 body of work.

Change in manuscript

No single new change in manuscript; see individual responses to lines 148ff, Figure 5, and line 209 for the specific additions.

Lines 151ff

Reviewer comment

Clarify which instruments were used to measure the variables on the horizontal axes in Fig. 3.

Response

We have added instrument identifications in the text and Fig. 3 caption: fluorescent particle concentrations (panel a) are from the WIBS; total particle mass (panel b) and organic and nitrate aerosol mass (panels d, e) are from the AMS; >0.5 µm particle number concentration (panel c) is from the aerodynamic particle sizer (APS); black carbon mass concentration (panel f) is from the aethalometer.

Change in manuscript

Added two instrument-identification parentheticals in Sect. 3.2 and corresponding instrument names in the Fig. 3 caption. In the body text, *"(fluorescent biological aerosol particles from the WIBS; total particle mass from the AMS; >0.5 µm number concentration from the APS)"* was inserted after *"horizontal axes"*, and *"(all three from the AMS except BC, which is from the aethalometer)"* was inserted after the list of bottom-panel variables.

Line 155 and Table 1

Reviewer comment

Previously, it is implied that organic mass and >500 nm concentration were included based on intuition and not a high-skill ranking.

Response

We acknowledge this apparent inconsistency and have made the dual selection rationale explicit: variables were chosen either because they ranked highly in the random forest analysis (fluorescent particles, BC, nitrate, particle mass) or because of their established physical connection to INP in prior parameterizations or site-specific studies (>0.5 µm concentration per DeMott/Tobo; organic mass per Hyytiälä biogenic seasonality). This is now stated clearly in Sect. 3.2.

Change in manuscript

Clarifying sentence added (see also response to line 144 above); no change to Table 1 entries.

Figure 3 and related interpretation

Reviewer comment

Explain why the data are split for the Pearson correlation analysis if the goal is to investigate the variables' predictive power for INP concentrations. The difference between PINC and PINCii data shows that the correlations are only good for a subset, not generally. Provide a discussion on what this implies for the many campaign-based INP correlations reported in the literature.

Response

We thank the referee for this important comment.

The PINC and PINCii data are plotted separately because the two instruments operated in distinctly different seasons—PINC in winter (February–April) and PINCii in spring/summer (April–June)—with distinct ambient meteorological and aerosol conditions. Pooling the two datasets would conflate these seasonal differences and conceal a key finding: the correlation between INP concentrations and the examined predictor variables is strongly season-dependent. The PINC winter data show essentially no correlation with any of the monitored variables (adjusted $R^2 \leq 0$), whereas the PINCii spring/summer data yield moderate correlations with several aerosol proxies (adjusted $R^2 \approx 0.5$ for fluorescent particles and nitrate). Merging the two would produce an artificial intermediate correlation coefficient that misrepresents the behaviour of each individual subset.

The referee correctly identifies the central implication: the correlations shown in Fig. 3 are specific to the PINCii spring/summer subset and do not hold generally across the full campaign. This is an important and candid finding that we have made more explicit in the revised manuscript.

This result has significant implications for campaign-based INP correlations reported in the literature. Many published INP–aerosol correlations are derived from intensive field campaigns of limited duration (days to weeks), often conducted during a single season or under specific meteorological regimes. Our results suggest that such correlations are inherently conditioned by the ambient aerosol and environmental conditions that prevail during the specific measurement period. A correlation identified during a spring campaign at a boreal forest site may not hold during winter at the same site, let alone at a different location or ecosystem. Authors and readers of campaign-derived INP correlations should therefore exercise caution when extrapolating or generalising such relationships beyond the specific conditions under which they are established.

Change in manuscript

Added to Sect. 3.2: *"Data for PINC and PINCii are shown separately because the two instruments operated in different seasons (winter vs. spring/summer) with distinct ambient aerosol and INP characteristics; combining them would conflate seasonal differences and obscure the key finding: the observed correlations are specific to the PINCii spring/summer subset and do not hold generally across the full campaign (the PINC winter data show no meaningful correlation with any of the examined predictors). [...] This subset-specific character of the correlations has broader implications for campaign-based INP correlations reported in the literature: INP–aerosol correlations derived from short, season-specific field campaigns are conditioned on the ambient regime of the measurement period and should be interpreted with caution when generalised to other seasons, sites, or ecosystems."*

Figure 3

Reviewer comment

For all panels, consider marking significant r values with an asterisk instead of reporting absurdly small p values. a), c): Check units on the horizontal axes. The different Pearson correlations for PINC and PINCii data, which seem to overlap for the most part in the log–log scatter plots, are surprising. Clarify if the correlation was calculated using the raw data or the log-transformed data. Consider showing the data on a linear scale. Provide the number of PINC and PINCii data points. Are they the same in each panel? Why are hourly mean data used here, and does using the mean affect the outcome of the analysis compared to using 20 min data?

Response

We thank the referee for these detailed technical comments on Fig. 3.

(a) **Asterisks for significance:** We have added asterisks to indicate statistically significant correlations ($p < 0.05$) and removed the previously reported exact p-values.

(b) **Units on horizontal axes:** We have verified and corrected the units on all horizontal axes in panels (a) and (c).

(c) **Differing Pearson correlations for apparently overlapping data:** The visually overlapping scatter in the log–log plots can yield different Pearson r values because the correlation is sensitive to the full range and distributional structure of the data, not just the visual density of points. The winter PINC data span a narrower and lower range of both INP and predictor values; thus the correlation structure is distinct even where the scatter appears to overlap with the spring/summer PINCii data.

(d) **Log-transformed vs. raw data:** The Pearson correlation coefficients were calculated from the raw (untransformed) data; the log–log axes are used solely for visualisation to span the wide dynamic range of the measurements. We have clarified this explicitly in the updated Fig. 3 caption.

(e) **Linear scale:** We considered adding linear-scale versions of the panels but, given the several orders of magnitude spanned by both INP concentrations and the predictor variables, a log–log representation is the most practical for displaying the full dataset.

(f) **Number of data points:** We have added the number of PINC and PINCii data points to the Fig. 3 caption. The number of points differs between panels because different sensors have different data availability owing to maintenance, calibration, and/or instrument downtime.

(g) **Hourly vs. 20-minute means:** We clarify in the caption that hourly means were used to align with the coarser temporal resolution of some complementary datasets; repeating the analysis using the native 20-minute data cadence yields qualitatively similar Pearson correlations.

Change in manuscript

Figure 3 updated with asterisks and verified units on horizontal axes. Caption expanded with: PINC and PINCii data point counts, clarification that Pearson coefficients were calculated from raw (untransformed) data with log–log axes used for visualisation only, rationale for use of hourly means, and confirmation that 20-minute data yields qualitatively similar results.

Lines 192ff

Reviewer comment

Eq. (2) reproduces Eq. (2) from Tobo et al., 2013, which uses $n_{AP>500\,\mathrm{nm}}$. However, the text seems to imply that Eq. (3) uses the number of fluorescent particles. How would the updated Eq. (3) from Tobo et al., 2013 perform? Which formula was used for Figure 8 in Brasseur et al., 2022?

Response

Both Eq. (1) (DeMott et al., 2010) and Eq. (2) (Tobo et al., 2013) in our manuscript use the number concentration of aerosol particles larger than 0.5 µm ($n_{AP>0.5\,\mu\mathrm{m}}$) as the predictor—not fluorescent particle counts. The Tobo et al. (2013) paper also provides a separate parameterisation using fluorescent biological aerosol particle (FBAP) concentrations (their Eq. 3), but we did not apply that variant here; our Eq. (2) follows the original $n_{AP>0.5\,\mu\mathrm{m}}$-based form. We have clarified this distinction explicitly in the text. Regarding Brasseur et al. (2022) Fig. 8: that figure used the same $n_{AP>0.5\,\mu\mathrm{m}}$-based Tobo (2013) formulation, consistent with our Eq. (2).

Applying the FBAP-based Tobo (2013) variant would require matching the FBAP dataset to the parameterisation's calibration dataset and would be an additional, substantial analysis.

Change in manuscript

Added the following note after Eq. (2): *"Note that both Eqs. (1) and (2) use the total aerosol number concentration $n_{\mathrm{AP}>0.5\,\mu m}$ as the predictor. Tobo et al. (2013) also provide a separate parameterisation based on fluorescent biological aerosol particle (FBAP) concentrations; that variant is not applied here."*

Figure 4

Reviewer comment

Would $p = 1$ not require $\chi^2 = 0$? Double-check the values. χ^2 for PINC data is huge and $p = 0$. The bimodal fit does not seem visually superior.

Response

We thank the reviewer for flagging these anomalies. Upon careful re-examination, both issues stemmed from the same underlying error in the statistical workflow.

(i) Anomalous $p \approx 1$ in panel (a) and the large χ^2 with $p = 0$ for PINC. The χ^2 goodness-of-fit test was erroneously applied to relative frequencies (dimensionless values summing to 1, each $\approx$ 0.01–0.10) rather than to raw bin counts. Because $\chi^2 = \sum(O - E)^2/E$, using small fractions systematically drove the statistic toward zero, producing artificially large p-values ($p \approx 1$) regardless of fit quality—explaining the anomaly flagged in panel (a). This has been corrected: the χ^2 test now operates on raw bin counts as observed frequencies, with expected counts derived by integrating the fitted probability density over each bin and scaling by the total sample size N. Degrees of freedom are adjusted for the number of estimated parameters (ddof = 2 for the unimodal fit; ddof = 5 for the bimodal fit).

(ii) Distorted bimodal fits in panel (b). A second error compounded the bimodal results: `curve_fit` was minimising residuals between the log-normal PDF (units: per unit x) and the relative frequencies (dimensionless), without accounting for the variable widths of the log-spaced bins. This mismatch distorted the optimised component widths and mode locations, particularly for the PINC dataset whose distribution is more broad. The bimodal fitting objective has been corrected to match probability masses (PDF $\times$ bin width) against relative frequencies. Initial parameter guesses were also made data-adaptive (placed at the 25th and 75th percentiles of each dataset) to reduce sensitivity to poor local minima.

With these corrections, the figure legends report $\chi^2 = 81.98$ ($p = 0.001$) and $\chi^2 = 157.19$ ($p < 0.001$) for the unimodal PINCii and PINC fits, respectively, so the strict unimodal log-normal null is rejected for both instruments at conventional significance. For the bimodal fits, PINCii attains $\chi^2 = 63.54$ ($p = 0.028$), i.e. a lower statistic than the unimodal case but still marginally inconsistent with the data at the 5% level, while the PINC bimodal fit remains rejected ($\chi^2 = 242.72$, $p < 0.001$) and does not improve overall agreement—consistent with the reviewer's visual assessment that the bimodal improvement was not compelling. The revised Figure 4 matches these values and is provided for comparison with the previous version.

Change in manuscript

- Updated the image for Figure 4 to the corrected version.
- Updated the Figure 4 caption to: *"The χ^2 goodness-of-fit test uses raw bin counts as observed frequencies, with expected counts derived by integrating the fitted probability density over each bin and scaling by total sample size N; degrees of freedom are adjusted for the number of estimated parameters (*ddof = 2 *for unimodal;* ddof = 5 *for bimodal fits). The unimodal fits yield $\chi^2 = 81.98$ ($p = 0.001$) for PINCii and $\chi^2 = 157.19$ ($p < 0.001$) for PINC. The bimodal fits give $\chi^2 = 63.54$ ($p = 0.028$) for PINCii and $\chi^2 = 242.72$ ($p < 0.001$) for PINC; χ^2 decreases for PINCii relative to the unimodal case, but the test still rejects the bimodal fit at the 5% level ($p = 0.028$), whereas the PINC bimodal model is not supported as an improvement over its unimodal counterpart."*
- Updated the body text discussing Figure 4 (Sect. 3.2) from *"From the fitting it is clear that to a large degree the observed INP concentrations are well represented by simple log-normal distributions. Moreover, the PINC and PINCii distributions are highly similar. The exception, manifest as a spike in the tail of the PINCii distribution, may be related to a particular source or series of events that were present during the PINCii sampling, but absent during the deep winter season. However, the divergence is not significant enough to robustly identify any real signal."* to *"From the χ^2 statistics (using raw bin counts; see caption), the formal tests reject the simple unimodal log-normal null at conventional significance for both PINCii ($\chi^2 = 81.98$, $p = 0.001$) and PINC ($\chi^2 = 157.19$, $p < 0.001$). Nevertheless, the histograms are broadly similar in shape and scale, so log-normality remains a useful approximate description of the bulk distribution. A modest elevated tail in the PINCii histogram motivates the bimodal decomposition in Fig. 4(b): for PINCii, χ^2 decreases to 63.54 ($p = 0.028$), whereas for PINC the bimodal fit does not improve overall agreement ($\chi^2 = 242.72$, $p < 0.001$). We therefore do not interpret the PINCii tail as evidence for a statistically robust separate population at the 5% level, and we caution*

that the χ^2 values indicate residual structure beyond ideal unimodal log-normal behaviour for both instruments."

Line 198

Reviewer comment

Explain why the CFDCs operating above water saturation were not measuring immersion freezing comparable to the parameterisation derived from datasets obtained with INSEKT. It would be interesting to see how well the parameterizations perform at low temperatures.

Response

CFDCs operate by exposing aerosol particles to a short traverse ($\sim$seconds) through a controlled cold, humid airstream, primarily activating condensation freezing and deposition nucleation, with limited potential for full immersion freezing. Bulk immersion freezing assays (such as INSEKT, used to derive the Schneider et al., 2021 and Brasseur et al., 2024 parameterisations) suspend particles in macroscopic water volumes cooled over minutes to hours, probing a fundamentally different nucleation pathway and timescale. Because the two approaches probe different physical processes, direct quantitative comparison is not straightforward. We have added a brief clarification to the text. Benchmarking the parameterisations at low temperatures would require dedicated datasets not available in the current study.

Change in manuscript

Added the following sentence to Sect. 3.3: *"Direct comparison is not straightforward because CFDCs primarily activate condensation and deposition freezing on a timescale of seconds, whereas the bulk immersion assays used to derive the Schneider et al. (2021) and Brasseur et al. (2024) parameterisations probe a different nucleation pathway over longer timescales."*

Figure 5

Reviewer comment

Double-check the calculation using Tobo et al., 2013. Figure 8 in Brasseur et al., 2022 shows good agreement between Tobo 2013 and PINC/PINCii measured INP concentrations during the instrument intercomparison.

Response

We have re-verified our implementation of the Tobo (2013) parameterisation and confirm the calculation is correct. We have carried out a careful, point-by-point comparison between our Fig. 5 and Brasseur et al. (2022) Fig. 8, and find that the apparent discrepancy is explained by three fundamental differences in dataset scope, comparison methodology, and aerosol conditions:

(i) Dataset scope. Brasseur et al. (2022) Fig. 8 covers only four targeted intercomparison days: 22 March, 28 March, 26 April, and 28 April 2018, selected specifically to maximise temporal overlap between instruments. Brasseur et al. (2022) themselves caution that the parameterisation comparison "might not be representative of the entire HyICE-2018 campaign." Our Fig. 5 covers the complete PINC deployment (19 February–2 April) and PINCii deployment (22 April–10 June), spanning the full range of seasonal aerosol and biological conditions at SMEAR II.

(ii) Qualified and limited agreement in Brasseur et al. (2022). Brasseur et al. (2022, their Fig. 11 and Sect. 3.2.3) consolidate all intercompared online and offline instruments for 28 March 2018 into a single INP temperature spectrum and count how many data points fall inside each parameterisation's shaded envelope (for Tobo (2013) and DeMott (2010), the envelope spans the APS $N_p(> 0.5\,\mu\text{m})$ daily mean ±1 standard deviation). They report 35% inside Tobo (2013), 3% inside DeMott (2010), and 19% inside Schneider et al. (2021)—so Tobo is "best" in a relative sense only. Their Fig. 8 time series is discussed qualitatively over the four intercomparison days; on 22 March—the only winter intercomparison day for PINC—they explicitly state that "none of the parameterizations successfully represents the measured concentrations," with measured INP concentrations both above and below the Tobo (2013) band at different times of day. The description of Tobo (2013) as showing "good agreement" is therefore relative, qualified, and does not hold for the winter portion of the campaign.

(iii) Comparison methodology. Brasseur et al. (2022) Fig. 8 presents a time-series visualisation in which measured INP concentrations are plotted alongside parameterisation *shaded bands* representing a −29 to −32°C temperature window. This is fundamentally different from our Fig. 5, which is a scatter plot of model-predicted versus measured INP concentrations evaluated by the coefficient of determination (R^2). Time-series visual overlap within order-of-magnitude bounds does not translate to predictive skill: even the consolidated Fig. 11 accounting of Brasseur et al. (2022) leaves most spectrum points outside the Tobo envelope (35% within), and such partial overlap will still yield strongly negative R^2 in a scatter plot if individual predicted and measured pairs are poorly correlated.

Beyond these methodological distinctions, the two PINCii intercomparison days (26 and 28 April) occurred immediately after the snowmelt transition at SMEAR II, when biological particle concentrations were increasing (Brasseur et al., 2022, their Fig. 6c). These two days represent the aerosol conditions most favourable to a parameterisation anchored on biological aerosol particles, and are not representative of the broader PINCii spring/summer measurement period when new particle formation (NPF) events frequently grow secondary organic aerosol particles into the >0.5 μm size range without a commensurate increase in biological INPs.

We have substantially expanded the manuscript discussion (Sect. 3.3) to reflect this nuanced comparison.

Change in manuscript

Replaced the single added sentence in Sect. 3.3 with: *"The better performance noted by Brasseur et al. (2022) (their Fig. 8 time series, and the consolidated 28 March inter-instrument spectrum summary in their Fig. 11) reflects a focused comparison over four targeted intercomparison days that are not representative of the full seasonal range covered here; Brasseur et al. (2022) themselves caution that their comparison 'might not be representative of the entire HyICE-2018 campaign.' Direct comparison between CFDC-based and bulk immersion freezing assay results is not straightforward because CFDCs primarily activate condensation and deposition freezing on a timescale of seconds, whereas bulk immersion assays (such as INSEKT, used to derive the Schneider et al. (2021) and Brasseur et al. (2024) parameterisations) probe a different nucleation pathway over longer timescales."*

Line 209

Reviewer comment

Clarify why in Figure 5b) Tobo 2013 never performs well.

Response

The systematic overprediction by Tobo (2013) for PINCii data across the entire spring/summer deployment is the result of two compounding factors that are specific to the Hyytiälä boreal environment during this period.

(i) Tobo (2013) calibration environment vs. SMEAR II spring aerosol. The Tobo (2013) parameterisation was developed at the Rocky Mountain Biological Laboratory (RMBL), a mid-latitude temperate ponderosa pine forest in Colorado, USA. At that site, particles larger than 0.5 μm are predominantly primary biological aerosol particles (fungal spores, plant debris, pollen fragments), and thus the high-$n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ conditions directly correspond to elevated biological INP concentrations. At SMEAR II during spring and early summer, new particle formation (NPF) events—for which Hyytiälä is globally renowned (Dal Maso et al., 2005; Kulmala et al., 2013)—nucleate and grow secondary organic aerosol (SOA) particles into the >0.5 μm size range over the course of hours to days. These SOA-grown particles inflate $n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ substantially without a commensurate increase in biological INP concentrations, because secondary organic material is a poor ice nucleant. The parameterisation therefore "sees" elevated aerosol particle counts and predicts large INP concentrations that are not realised.

(ii) Sub-Arctic boreal ecosystem vs. temperate forest. The boreal forest of southern Finland in spring and early summer is ecologically distinct from a temperate mid-latitude forest. Biological primary aerosol emission is lower in the early season (April–June) than in the peak summer of a temperate forest, and the diversity of biogenic INP sources is correspondingly reduced. Accordingly, the fraction of $n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ that consists of INP-active biological material is lower at Hyytiälä than the RMBL calibration environment, causing the parameterisation to systematically overestimate.

We note that this mechanism is consistent with the observation in Brasseur et al. (2022) that Tobo (2013) performs relatively better on 26 and 28 April—immediately after snowmelt when biological fractions in the aerosol were rising (their Fig. 6c)—than over the broader spring/summer PINCii period. We have updated the manuscript text to reflect this mechanistic explanation.

Change in manuscript

Added after Table 1 in Sect. 3.3: *"The persistent overprediction by Tobo et al. (2013) for PINCii reflects two site-specific factors: (i) at Hyytiälä, frequent NPF events grow secondary organic aerosol into the >0.5 µm size range (Dal Maso et al., 2005; Kulmala et al., 2013), inflating* $n_{\mathrm{AP}>0.5\,\mu m}$ *without a commensurate increase in biological INPs, whereas the parameterisation was calibrated at a North American temperate forest where large particles are predominantly primary biological aerosol; and (ii) the sub-Arctic boreal spring supports lower primary biological aerosol emission than the temperate calibration environment, further reducing the fraction of large particles that are ice-active at these temperatures."*

Line 218

Reviewer comment

Please provide a more in-depth interpretation of the coefficients found in Table 1.

Response

We have added a paragraph after Table 1 interpreting the fitted exponents and pre-factors.

Change in manuscript

Added the following paragraph after Table 1:

> *"The exponent j reflects the sensitivity of INP concentration to changes in the predictor: values near unity indicate approximately linear relationships (>0.5 µm number: $j = 1.13$; BC mass: $j = 1.03$), values well below unity suggest weak sensitivity (organic mass: $j = 0.56$), and near-zero or negative values indicate absence of a meaningful relationship (all PINC predictors). The pre-factor i sets the absolute scale and is influenced by the ambient INP concentration range during the respective measurement period. The consistently low or negative adjusted R^2 for PINC confirms the absence of predictive skill during the winter period, regardless of the predictor chosen."*

Figure 6

Reviewer comment

Check units in panel titles for panels a and c.

Response

We have checked and corrected the units in panels (a) and (c) of Fig. 6.

Change in manuscript

Figure 6 panel titles updated with correct units.

Technical corrections

Line 36 ("bridging") Changed *"bridging"* to *"extending"* — see also the specific comment response to lines 148ff above.

Lines 132–134 (repetition) Removed — see response above.

Line 139 ("ice nucleation activity") Both occurrences replaced with "INP concentration" — see response above.

Line 162 ("to be ice active") Changed *"be ice active"* to *"contain INPs"*.

0.5 µm vs. 500 nm All occurrences standardised to “0.5 µm” throughout text, figure captions, and Table 1.

Brasseur et al., 2022 author list The author list in the bibliography has been updated to include the full list as also indicated by the referee: Brasseur, Z., Castarède, D., Thomson, E. S., Adams, M. P., Drossaart van Dusseldorp, S., Heikkilä, P., Korhonen, K., Lampilahti, J., Paramonov, M., Schneider, J., Vogel, F., Wu, Y., Abbatt, J. P. D., Atanasova, N. S., Bamford, D. H., Bertozzi, B., Boyer, M., Brus, D., Daily, M. I., Fösig, R., Gute, E., Harrison, A. D., Hietala, P., Höhler, K., Kanji, Z. A., Keskinen, J., Lacher, L., Lampimäki, M., Levula, J., Manninen, A., Nadolny, J., Peltola, M., Porter, G. C. E., Poutanen, P., Proske, U., Schorr, T., Silas Umo, N., Stenszky, J., Virtanen, A., Moisseev, D., Kulmala, M., Murray, B. J., Petäjä, T., Möhler, O., and Duplissy, J.

We thank the referee once more for the thorough and constructive review. We believe the revisions substantially improve the manuscript’s clarity, scientific framing, and accessibility.

Author Reply to Reviewer Comments

RC2 Reply

Dear Referee,

We thank you for the thoughtful and constructive review. Your comments address fundamental questions about the scope, framing, and interpretation of our study, and have prompted us to substantially improve both the scientific rigour and the clarity of the manuscript. We address each comment below in the order presented.

Where relevant, we note how your comments relate to changes that are also addressed in the companion RC1 reply, to ensure full consistency across both replies.

Major points

Screening of variables: methodology

Reviewer comment

(1) While you mention their names around l. 129, any more methodological information is absent. Please elaborate on that in a separate methodology section.

Response

We agree that the ML methodology required more explicit documentation, and this comment is consistent with feedback from Reviewer 1 (see the companion RC1 reply). We have added a dedicated paragraph to Sect. 2.2 that describes the algorithms applied (random forest and decision tree models for feature importance ranking; principal component analysis and K-means clustering for dimensionality reduction), clarifies that no train/test data split was performed (the purpose was exploratory importance ranking), and explains the time harmonization procedure. We have also added a closing sentence to Sect. 2.2 restating that these methods were applied as exploratory tools, not to build or evaluate a predictive model.

Change in manuscript

See the companion RC1 reply for the corresponding manuscript change (modifications to Sect. 2.2).

Reviewer comment

(2) Where can we see that the separate treatments resulted in qualitatively similar results (l. 131 and l. 159)? Fig. 2 only shows one of these, and a comparison should at least be part of the Appendix.

Response

The referee suggests a good point. However, although qualitatively similar, the separate treatments were not robust and small changes, for example in the random seed were observed to lead to differences. Thus, presenting any fixed set of results for comparison would be somewhat arbitrary. With this in mind we prefer to present only one “qualitative" example, while we try to emphasize the *qualitative* nature of the result.

Duplicate text describing the variable exclusion criteria (Sect. 3.2, l. 132–134) has been removed to reduce repetition. Additionally, we modify the text at l. 131 to read “which are illustrated in Fig. 2” (replacing “as illustrated”), making clear that Fig. 2 shows the results from one representative method (random forest). We add a sentence acknowledging that a formal multi-method comparison figure would arbitrarily relate results from the methodologies in an artificial way.

Change in manuscript

Changed “as illustrated in Fig. 2” to “which are illustrated in Fig. 2”. Added: *“A formal multi-method comparison (e.g., side-by-side feature rankings from random forest, decision tree, and PCA) is not shown but was verified by the authors during analysis; these approaches consistently highlighted similar top-ranking variables.”*

Reviewer comment

(3) How was the subset (of the 84 variables that you analysed) selected for Fig. 2? Are these the ones with the highest importance?

Response

Yes. Figure 2 shows the top-ranked variables by feature importance from the random forest analysis, ordered by decreasing importance score. We have clarified this in the updated Fig. 2 caption.

Change in manuscript

Added to the Fig. 2 caption: *“Variables are shown in decreasing order of random-forest importance. The six variables examined in pairwise correlation in Fig. 3—fluorescent particle number concentration (WIBS), particle mass (1 nm–10 μm aerodynamic diameter, AMS), number concentration of particles with diameters >0.5 μm (APS), organic aerosol mass concentration (AMS), nitrate aerosol mass concentration (AMS), and black carbon mass concentration (aethalometer)—appear among the top-ranked variables in this panel (selection criteria in Sect. 2.2).”*

Uptake of screened variables

Reviewer comment

From the variables in Fig. 2, you make another selection for Fig. 3. (l. 144) states that these were selected "based upon their highly ranked outcomes and physical intuition." On a first glance, this appears a reasonable and often-used approach, but in the context of your study it seems to invalidate the idea that the study is based on: screening as many variables as possible. If you exclude those variables that you didn't expect to show a correlation (physical intuition) after the screening, why include them in the first place?

Many of the variables that appear with high importance in Fig. 2 also remain unmentioned — I think at the least you ought to explain why you exclude them from further analysis (perhaps in a Table?).

Response

We appreciate this critical observation. The referee correctly identifies a tension between the open screening philosophy and the subsequent use of physical intuition as a filter. We wish to clarify that the screening was genuinely open-ended: all 84 variables entered the random forest analysis without prior filtering by physical expectation. The subsequent selection of six variables for detailed scatter-plot examination in Fig. 3 does not invalidate the screening; rather, it reflects the practical constraint that a detailed correlation analysis with visualisation is feasible only for a limited number of variables. The screening's primary value is in revealing which variables rank highly, including *unexpected* ones (e.g., black carbon, nitrate), which we do investigate and discuss.

We have replaced the ambiguous "physical intuition" phrasing with an explicit dual-selection justification: *"The six variables shown in Fig. 3 were selected on two complementary grounds: (i) high importance rank in the random forest analysis (fluorescent particles, black carbon, nitrate, particle mass); and (ii) established physical or empirical connection to INP in prior literature or at this specific site."* This makes the rationale transparent.

Regarding other high-ranking variables, we add a statement acknowledging that other highly ranked variables (e.g., acetone, methanol) largely co-vary with the selected variables (particularly nitrate) and thus provide redundant information. Our selection therefore covers the principal top-ranked predictors after accounting for redundancy among co-varying variables, and it is unlikely that lower-ranked variables would yield additional insight.

Change in manuscript

- The dual-selection justification is added (see the companion RC1 reply, "Line 144" response).
- Added: *"Several other highly ranked variables (e.g., acetone, methanol concentrations) are not examined individually because they largely co-vary with the selected predictors (particularly nitrate) and thus provide redundant information for the purposes of this exploratory analysis."*

Title and framing

Reviewer comment

The title suggests that you used machine learning extensively and that you are predicting INP properties. In my reading, this does not reflect what you did and what is explained in the manuscript:

- the ML was only used to screen variables and not documented well (see above)
- your modified parameterisation shows only modest skill, and only for spring measurements. It is also only a small part of the study.
- The only thing you are predicting are INP concentrations, thus only one INP property.

This can easily be read as overselling [. . .] As you state in the conclusion, "no single parameter emerges that is strongly linked to INP." The paper title and framing need to reflect this.

Response

We agree with the referee's assessment (see also the companion RC1 reply). We change the title from "Predicting" to "Exploring" and add explicit statements that the ML methods were used as importance-ranking tools, not to build a predictive model.

However, we acknowledge that the referee's concern goes further: even "Exploring" may suggest that ML is the centerpiece of the study, whereas the main scientific message is the *null result* — that no single parameter strongly predicts INP concentrations, even with extensive co-located measurements.

Therefore, we further refine the title to make plain the null result, and finally suggest: *"Exploring Ice Nucleation Particle concentrations in a Boreal Environment: limits of machine-learning-assisted variable screening."*

This addresses the referee's concern by:

1. Replacing "properties" with "concentrations" (only one INP property is examined).
2. Adding "limits of. . . variable screening" to signal the null result directly in the title.
3. Retaining "Exploring" inspired by the RC1 comments.

We have also replaced "INP variability" with "INP concentrations" throughout the abstract and text where appropriate, following the referee's observation that variability per se is not analyzed.

Change in manuscript

- Title: changed from *"Exploring Ice Nucleation Particle properties in a Boreal Environment using machine learning"* to *"Exploring Ice Nucleation Particle concentrations in a Boreal Environment: limits of machine-learning-assisted variable screening."*
- "INP variability" replaced with "INP concentrations" in the abstract and where applicable.

Conclusion from the null results

Reviewer comment

As I stated above, I think the "null result" is an important one. However, the conclusion jumps from "drawing strong conclusions that can illuminate causality will likely remain illusive" to suggesting more, longer and heavily-equipped measurements. [. . .] given your sobering findings, wouldn't the opposite make more sense? More and more measurements will NOT magically "help the community to identify key predictor parameters". Why do you think you couldn't find a correlation? What does this imply about the nature of INPs? What could one/the field do differently? [. . .]

Response

We thank the referee for this challenging and thought-provoking comment. The referee is right that the standard call for "more measurements" is insufficient — and potentially misleading — in the light of our null result. We have given this careful consideration.

The weak correlations found in our study, combined with the log-normal INP concentration distributions (consistent with random mixing and dilution of trace species; Ott, 1990), strongly suggest that INP at this site are dominated by long-range transport of aerosol from diverse, distant sources. In such a scenario, no local measurement suite — however comprehensive — can be expected to yield strong, causal predictor–INP relationships, because the INP identity and source vary stochastically with air-mass origin rather than responding to locally measured tracers. This interpretation is independently supported by Paramonov et al. (2020), who used the same PINC instrument and dataset at the same site and concluded: *"No single dominant local or regional source of INPs in the boreal environment of southern Finland could be identified. Rather, it is hypothesised that the INPs detected at SMEAR II are a result of long-range transport and dilution of INPs sourced far from the measurement site."*

We have added a new paragraph to the conclusions that directly engages with the referee's questions:

1. We acknowledge that the null result suggests a fundamental limitation: if INP are dominated by long-range transport, then local correlations will always be weak.
2. We propose that future campaigns should *complement* high-frequency INP measurements with source-apportionment tools (back-trajectory analysis, receptor modelling) to evaluate consistency with the transport hypothesis and constrain plausible source regions—rather than simply accumulating more of the same local measurements.
3. We note that the question "why can't we predict INP?" is itself a scientifically valuable finding that deserves explicit investigation.

We have preserved the existing recommendation for continued measurements at equipped stations (see also the companion RC1 reply) but placed it in the context of complementary analytical approaches.

Change in manuscript

The following paragraph is added to the conclusions:

> *"The weak correlations observed in this study, combined with the log-normal INP concentration distributions consistent with random dilution of trace species, suggest that INP at this boreal site are dominated by long-range transport from diverse, distant sources.*

In such a regime, no local measurement suite—however comprehensive—can be expected to yield strong, causal predictor–INP relationships, because the INP identity and source vary stochastically with air-mass origin. This interpretation is independently supported by Paramonov et al. (2020), who reached a similar conclusion using the same PINC dataset at the same site. We therefore propose that future campaigns should complement high-frequency INP measurements with source-apportionment tools (e.g., back-trajectory analysis, receptor modelling) to evaluate consistency with the long-range transport hypothesis and to constrain the geographic origins and aerosol types that plausibly contribute to the INP population—recognising that trajectory-based tools alone cannot uniquely verify INP provenance without tracer or receptor constraints. The question 'why can't we predict INP from local measurements?' is itself a scientifically valuable finding: it constrains the problem space and guides the field toward approaches that account for air-mass history rather than relying solely on local co-located tracers."

New parameterization does not respond to predictors

Reviewer comment

l. 222: "PINC measurements ... have almost no response to the predictors": it's the predicted INP that don't vary, while the measured INP do. However, I don't understand how this can be the case? Seeing the simple parameterisation formula, it implies that the predictors themselves don't vary. Is that correct? Fig. 1 implies otherwise, at least the particle mass is varying quite a bit there, so I do not understand why a parameterisation based on it would give constant results.

Response

We thank the referee for identifying this confusing passage. The issue is a combination of the wording and the mathematical properties of the fitted power law.

The predictors (e.g., particle mass) *do* vary during the PINC winter period, as visible in Fig. 1. However, the fitted exponent j in Eq. (3) for the PINC data is near zero or slightly negative for all six predictors (Table 1): particle mass: $j = -0.10$; organics: $j = 0.00$; fluorescent particles: $j = -0.05$; BC: $j = 0.03$, etc. In a power-law relationship $n_{\mathrm{INP}} = i \times X^j$, an exponent near zero means $X^j \approx 1$ regardless of the value of X, so the predicted INP concentration collapses to approximately the pre-factor i (i.e., the mean INP concentration), producing a near-constant prediction despite substantial predictor variability.

This is a direct consequence of the absence of a meaningful statistical relationship between the predictor and INP during winter: the best-fit power law effectively "gives up" by setting the exponent to zero, reverting to the sample mean. This is consistent with the negative adjusted R^2 values reported in Table 1 for all PINC predictors.

We have clarified the wording and added an explanatory sentence.

Change in manuscript

- Changed *"have almost no response to the predictors"* to *"yield near-constant predicted INP concentrations because the fitted exponents (j in Eq. 3) are near zero for all PINC predictors*

(Table 1), meaning the predicted values collapse to approximately the pre-factor i regardless of predictor variability."

- Added an explanatory paragraph to the table discussion: *"The exponent j reflects the sensitivity of INP concentration to changes in the predictor: values near unity indicate approximately linear relationships (. . .), values well below unity suggest weak sensitivity (. . .), and near-zero or negative values indicate absence of a meaningful relationship (all PINC predictors). The consistently low or negative adjusted R^2 for PINC confirms the absence of predictive skill during the winter period, regardless of the predictor chosen."*

Minor points

Abstract: 84 vs. 500 variables (l. 7)

Reviewer comment

[. . .] many of the variables that could have potentially been explored were excluded due to excessive NaN values or little variability, which left you with 84 variables. I think this is the number that should be mentioned in the abstract, rather than a grand 500 variables (l. 7).

Response

We agree that the abstract should be precise about the number of variables actually analysed. We now mention both numbers in the abstract: the initial scope (more than 500 variables monitored at SMEAR II) and the analysis set (84 retained after quality screening). This preserves the context about data richness while being transparent about the analysis scope.

Change in manuscript

Changed abstract from *". . . more than 500 high-resolution atmospheric, aerosol, and ecosystem variables measured continuously at Station for Measuring Ecosystem-Atmosphere Relations (SMEAR) II."* to *". . . more than 500 high-resolution atmospheric, aerosol, and ecosystem variables measured continuously at the Station for Measuring Ecosystem-Atmosphere Relations (SMEAR) II, of which 84 were retained after quality screening for the analysis."*

"INP concentrations" vs. "INP variability" (l. 7 and throughout)

Reviewer comment

Aren't INP concentrations what you're after? I don't see you using a measure of variability anywhere, nor do you correlate with variability itself.

Response

The referee is correct. Our analysis correlates INP concentrations (not their variability per se) with predictor variables. We have replaced "INP variability" with "INP concentrations" in the abstract and throughout the text where pertinent. See also the companion RC1 reply.

Change in manuscript

Replaced "INP variability" with "INP concentrations" in the abstract and in relevant passages throughout the manuscript.

Introduction

Line 18

Reviewer comment

They don't need to "form" below 0°C, the cloud forms as the cloud droplet forms.

Response

We agree that the original phrasing was imprecise. Mixed-phase clouds are defined by the co-existence of liquid water and ice at temperatures below 0°C; the cloud itself forms when droplets activate on CCN, and ice subsequently appears via heterogeneous nucleation. We have corrected the wording.

Change in manuscript

Changed *"for ice to exist, the temperature must be less than 0°C, meaning that at these temperatures liquid water would prefer to be ice"* to *"for ice crystals to form via heterogeneous nucleation, temperatures below 0°C are required; at such temperatures, ice is the thermodynamically stable phase, yet supercooled liquid water persists because a high kinetic barrier inhibits spontaneous freezing."* This also addresses the "school language" and "missing link" concerns below.

Line 20: "would prefer to be ice" and missing link

Reviewer comment

"Would prefer to be ice" — "school language". Also, the link between the sentences is missing: water would "prefer to be ice", yet it is not!

Response

The sentence containing "unstable co-existence of ice and liquid water" is replaced with a more precise thermodynamic description (see also the companion RC1 reply) ("thermodynamically metastable co-existence of ice and supercooled liquid water ..."). Moreover, we further refine the passage to remove the informal language, to provide the missing logical link (why liquid persists despite thermodynamic instability), and to integrate the explanation into a single coherent sentence.

Change in manuscript

The passage is now: *"Even with this inherently thermodynamically metastable co-existence of ice and supercooled liquid water – ice is the stable phase below 0°C, but a high kinetic barrier to nucleation sustains liquid water in the absence of efficient INPs – mixed-phase clouds are widespread ... ".*

Line 23: “underlying”

Reviewer comment

What is meant with “underlying”?

Response

The revision specifies the feedbacks and removes the ambiguous “underlying” (see also the companion RC1 reply). “especially in the Arctic where underlying feedbacks have amplifying effects” is replaced with *“where amplifying feedbacks—such as the ice-albedo feedback and the water vapour–temperature amplification characteristic of polar warming—have particularly pronounced effects.”*

Change in manuscript

See the companion RC1 reply for the corresponding manuscript change.

Line 24: “only”

Reviewer comment

“Only” — ice formation or ice in MPCs can also occur because of ice sedimenting into the cloud from another cloud above.

Response

We agree. Ice crystals can also enter a mixed-phase cloud via sedimentation from overlying ice clouds (e.g., “seeder–feeder” mechanism). We have added a brief parenthetical acknowledging this pathway.

Change in manuscript

Changed *“mixed-phase clouds only appear with the help of small particles”* to *“mixed-phase clouds primarily form with the help of small particles in the atmospheric aerosol”* and added a parenthetical: *“(ice can also enter a cloud via sedimentation from above, e.g., the seeder–feeder mechanism)”*.

Line 34: ICOS, ACTRIS, SITES citations and acronyms

Reviewer comment

Give citations for ICOS, ACTRIS and SITES (and explain the acronyms).

Response

We have added the full names and appropriate citations for all three infrastructure networks.

Change in manuscript

Changed *"e.g., ICOS, ACTRIS, SITES, etc."* to *"e.g., the Integrated Carbon Observation System (ICOS; Heiskanen et al., 2022), the Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS; Pandolfi et al., 2018), and the Swedish Infrastructure for Ecosystem Science (SITES; 2021)."*

Line 35: take-aways from previous campaign studies

Reviewer comment

What were the most important take-aways (about INPs) from these other campaign studies?

Response

We have expanded our connections to all HyICE publications, which include Paramonov et al. (2020), Schneider et al. (2021), Brasseur et al. (2022, 2024), and Vogel et al. (2024). We add a sentence summarizing the key findings from these prior publications.

Change in manuscript

Added after the list of HyICE citations: *"Key findings include the characterization of condensation/immersion mode INP concentrations and their likely origin from long-range transport (Paramonov et al., 2020), the identification of seasonal trends linked to biogenic emissions (Schneider et al., 2021), a complete campaign overview and INP instrument intercomparison (Brasseur et al., 2022), a description of flight-based measurements that include vertically resolved INP concentrations over the boreal forest (Brasseur et al., 2024), and the role of fluorescent biological aerosol particles as INP tracers (Vogel et al., 2024)."*

Line 40: "INP data" — be specific

Reviewer comment

"INP data": as for "variability" above, be specific.

Response

Consistent with the change from "ice nucleation activity" to "INP concentration" and with the referee's earlier comment about "variability," we have changed "INP data" to "INP concentrations."

Change in manuscript

Changed *"INP data"* to *"INP concentrations"*.

Line 41: "even" → "especially" and spurious correlations

Reviewer comment

"Even": I would argue that "especially" fits better, because the more variables you try (and the less physical intuition you have for them, see above), the more likely it is that you find spurious correlations. This is a wider issue that I think could be addressed or at least mentioned somewhere in the manuscript.

Response

The referee makes an excellent point. With 84 variables and limited sample sizes, the risk of finding spurious correlations is real. Paramonov et al. (2020) faced the same concern from their reviewers when correlating INP with multiple aerosol properties at the same site, and addressed it by adding explicit caveats (e.g., "correlation of [INP] with fluorescent particle concentration does not imply that INPs are necessarily fluorescent"). We have replaced "even" with "especially" and added a sentence about the multiple-comparison problem.

Change in manuscript

- Changed "even" to "especially."
- Added: *"When many variables are screened simultaneously, the probability of finding spurious correlations increases; the results presented here should therefore be interpreted as hypothesis-generating rather than hypothesis-confirming, and the identified associations require independent validation."*

Missing: physical intuition background in introduction

Reviewer comment

Since you draw on "physical intuition" to argue for the variables that you include in Fig. 3, I miss background on that intuition in the introduction. What have INPs been correlated with before?

Response

We augment the introduction to provide additional context for the reader with respect to aerosol properties that have previously been linked to INP concentrations. This background motivates both the screening approach and the subsequent variable selection. We have added a brief paragraph to the introduction.

Change in manuscript

Added text: *"Previous studies have linked INP concentrations to a range of aerosol properties, including total particle number concentration above 0.5 μm (DeMott et al., 2010; Tobo et al., 2013), fluorescent biological aerosol particle concentrations (Tobo et al., 2013), mineral dust loading (DeMott et al., 2015), and black carbon (DeMott, 1990). At the same site, Paramonov et al. (2020) found that no single parameter predicted INP concentrations over the full campaign period, although short-timescale correlations with black carbon, supermicron biological particles, and sub-0.1 μm particles* *were observed. These precedents motivate the present study's open-ended screening approach while providing the physical context for interpreting the resulting variable rankings."*

Methods and Data

Line 50: "two" but list 3

Reviewer comment

You say "two" but list 3.

Response

We thank the referee for this close read. The numeral "two" is correct: PINC and PINCii are the two continuous-flow diffusion chambers (CFDCs) deployed during HyICE-2018. The abstract wording "two … chambers, Portable Ice Nucleation Chamber I and II (PINC and PINCii)" could be parsed as three separate items (chamber type, full names, abbreviations) rather than as two instrument names; we have revised the sentence to make the pairing explicit.

Change in manuscript

Abstract: replaced the ambiguous appositive with *"Two continuous-flow diffusion chambers (CFDCs), PINC and PINCii (Portable Ice Nucleation Chambers I and II), were deployed with high-frequency sampling to measure INP concentrations"*, so "two" unambiguously refers to the two instruments. The Methods text already states that *"two continuous flow diffusion chambers (CFDCs; PINC … and PINCii …) … were operated"* (Sect. 2.1); no change was required there.

Lines 53–58 and 73–74: repetition

Reviewer comment

No need to repeat the introduction or move this content there if it hasn't been covered.

Response

We agree that these passages duplicate introductory material. We have consolidated the text by removing the repetition from the Methods section and ensuring the relevant content appears only once.

Change in manuscript

Removed duplicate introductory material from Sect. 2.2 (the first two sentences restating the motivation for co-locating measurements at SMEAR II) and from Sect. 2.3 (the opening sentence restating the study's overarching objective). The relevant scientific context is retained in the Introduction.

Line 59: singular "purpose"

Reviewer comment

Singular purpose.

Response

Corrected.

Change in manuscript

Changed *"purposes"* to *"purpose"*.

Line 68: instrument documentation

Reviewer comment

In principle I think all data that you use should be described with this much information, or else please describe why you describe only this one and where information on the others can be found.

Response

We have added a sentence to Sect. 2.2 directing readers to Brasseur et al. (2022) for HyICE-2018 instrument specifications and to the SmartSMEAR portal for the full SMEAR II monitoring suite metadata (see also the companion RC1 reply). Given that we use over 80 variables from dozens of instruments, a complete instrument table in the manuscript would be impractical; instead, we provide these comprehensive references.

Change in manuscript

See the companion RC1 reply for the corresponding manuscript change.

Line 104: main take-aways from comparison

Reviewer comment

Please state the main take-aways from the comparison.

Response

We have added a brief summary of the key findings from the PINC/PINCii comparison described in Brasseur et al. (2022).

Change in manuscript

Added: *"The comparison showed that PINC and PINCii agree within a factor of two during the overlapping measurement period, with both instruments capturing the same general trends in INP* *concentrations; minor differences in absolute concentration are attributable to the 1 K difference in operating temperature ($T_l = -31\,^{\circ}C$ for PINC, $-32\,^{\circ}C$ for PINCii) and instrument-specific detection efficiencies."*

Lines 110–111: "slightly greater variability"

Reviewer comment

What do you mean with "slightly greater variability"? Can you quantify this, for example by giving a range or fitting a distribution?

Response

We did not intend a formal distribution fit or numeric summary (e.g., geometric standard deviation or interquartile range in concentration decades). The remark referred to a qualitative visual impression from Fig. 1(a): the PINCii spring and early summer segment shows a somewhat wider envelope on the logarithmic concentration axis than the PINC winter segment. Because PINC and PINCii did not operate simultaneously, we do not report paired distribution statistics between instruments here; such numbers would not resolve the referee's question in a physically interpretable way without a dedicated intercomparison design. We have revised the sentence to state explicitly that the comparison is descriptive (visual inspection of Fig. 1(a)) and to link the pattern to expected seasonal differences in aerosol sources and meteorology, without implying that variability has been quantified.

Change in manuscript

Changed *"the PINCii measurements in spring and early summer appear to contain slightly greater variability"* to *"by visual inspection of Fig. 1(a), the PINCii spring and early summer time series appears somewhat more spread on the logarithmic concentration axis than the PINC winter series, consistent with seasonal changes in aerosol sources and meteorology; we do not report distribution fits or geometric moments here because PINC and PINCii did not operate simultaneously, so we restrict the comparison to this qualitative description."*

Results

Figure 1 caption: PINCii squares and missing FBAP data

Reviewer comment

I don't see squares for PINCii in (a)? Why are there no measurements of fluorescent biological aerosol particles before 03-15 in (c)?

Response

We have added measurement temperature information and corrected units in the Fig. 1 caption (see also the companion RC1 reply). Regarding the referee's additional questions:

(1) PINC and PINCii data are distinguished by colour in panel (a): PINC is plotted in blue and PINCii in red. We have added a note in the caption clarifying the colour coding and measurement temperatures.

(2) The WIBS instrument (measuring fluorescent biological aerosol particles) was deployed starting 11 March 2018, approximately one month into the campaign. Prior to that date, no FBAP data are available. We have added an explanatory note to the caption.

Change in manuscript

Updated Fig. 1 caption: (1) Added temperature and colour clarification: *"INP concentrations were measured at $T_l = -31\,^{\circ}C$ (PINC; blue) and $T_l = -32\,^{\circ}C$ (PINCii; red)."*; (2) Added: *"WIBS fluorescence data begin on 11 March 2018, when the instrument was first deployed."*

Lines 132–137: repetition

Reviewer comment

Repetition or move to Methods if it hasn't been said before.

Response

See the companion RC1 reply for the corresponding manuscript change.

Line 142: investigate surprising variables more closely

Reviewer comment

Related to the major point above, why don't you investigate the link between surprising variables and INP concentrations more closely? This would add value to your idea of sampling as many variables as possible.

Response

We agree that deeper investigation of surprising high-ranking variables is valuable, and in fact this influenced our choices throughout the manuscript. Perhaps, our most surprising finding is the high importance of black carbon (BC), which we do investigate and discuss in detail – including the independent support from Paramonov et al. (2020) who also found BC–INP correlation (their Fig. 5). We have added discussion of possible mechanisms (aging, oxidation, coating with organic material) and cited relevant literature.

Other (perhaps) surprising variables (e.g., acetone, methanol) largely co-vary with nitrate and are discussed as redundant air-mass tracers.

Change in manuscript

Added: *"A systematic investigation of all high-ranking but unexpected variables is an important avenue for future work."*

Line 145: co-variance not stated as exclusion reason

Reviewer comment

You give co-variance as an example for exclusion, but do not state it as a reason for exclusion in the sentence prior.

Response

We have corrected this logical gap by adding co-variance as an explicit exclusion criterion in the sentence preceding the example.

Change in manuscript

The existing example ("acetone and methanol, which largely co-vary with nitrate, are not [examined]") already demonstrates co-variance as an exclusion rationale. As addressed in the companion RC1 reply, the dual-selection justification now provides the explicit selection criteria (high random-forest importance and established physical/empirical connection), making co-variance exclusion implicit in the "redundant information" phrasing.

Figure 2: cross-reference to Fig. 3; variable specification

Reviewer comment

The comparison to Fig. 3 is difficult. Could you highlight the variables that you chose to investigate in detail in Fig. 3? Also, many of these variables need further specification, for example "nitrate" what? Number concentration?

Response

We address the comparison by spelling out, in the Fig. 2 caption, the same six measured quantities and instruments as in Fig. 3 (WIBS fluorescence; AMS particle mass, organics, and nitrate; APS number concentration >0.5 μm; aethalometer black carbon). Readers can therefore locate the corresponding bars by name and see explicitly whether "nitrate" refers to aerosol mass from the AMS, etc. The y-axis labels on the published figure file remain the short SMEAR II metadata names; full quantities are given in the caption.

Change in manuscript

Fig. 2 caption: added explicit listing of the six Fig. 3 variables with instrument and quantity: *"The six variables examined in pairwise correlation in Fig. 3—fluorescent particle number concentration (WIBS), particle mass (1 nm–10 μm aerodynamic diameter, AMS), number concentration of particles with diameters >0.5 μm (APS), organic aerosol mass concentration (AMS), nitrate aerosol mass concentration (AMS), and black carbon mass concentration (aethalometer)—appear among the top-ranked variables in this panel (selection criteria in Sect. 2.2)."*

Figure 2 caption: decision tree explanation

Reviewer comment

The explanation of the decision tree belongs into the Methods section (and, as stated above, the other ML methods need to be explained there as well).

Response

We agree. A dedicated ML methodology paragraph is added to Sect. 2.2. We have moved the decision tree explanation from the Fig. 2 caption to this methods paragraph and simplified the Fig. 2 caption to reference Sect. 2.2 for methodological details.

Change in manuscript

Removed the Random Forest ensemble explanation from the Fig. 2 caption (that material is covered in Sect. 2.2). The sentence on bar height now reads *"Higher values indicate a stronger statistical association with INP concentrations"* (in place of the longer decision-tree wording), and the caption adds the cross-reference listing the six Fig. 3 variables with full specifications (as under comment (3) and the Fig. 2 comparison response above), without bold formatting in the figure.

Lines 152–156: repetitive

Reviewer comment

Repetitive of what's been said just before.

Response

We have removed the repetitive passage.

Change in manuscript

Removed the repetitive text at l. 152–156.

Figure 3: PINC = winter, PINCii = spring

Reviewer comment

You may want to remind the reader that PINC corresponds to winter and PINCii to spring measurements.

Response

We have added extensive text explaining the data splitting (see also the companion RC1 reply). We also add a brief reminder directly in the Fig. 3 caption.

Change in manuscript

Added to Fig. 3 caption: *"PINC data (blue) correspond to winter measurements (19 February–2 April); PINCii data (red) to spring/summer (22 April–10 June)."*

Line 160: correlations $\neq$ causation

Reviewer comment

Why would a method that shows you correlations be enough to "shed light on causation"? Please explain what made you reason that the correlations you find do not reflect causations.

Response

The referee is correct that our correlation-based approach cannot establish causation. The original phrasing was misleading. We have clarified that the ML and correlation analysis were used to identify statistical associations, not causal relationships. The reasons we believe these correlations are not causal include: (i) the well-known rarity of INPs relative to the total aerosol population, meaning that any observed bulk-aerosol correlation likely reflects co-varying air-mass properties rather than direct INP composition; (ii) the absence of single-particle analysis of ice crystal residuals in our setup, which would be needed to directly identify INP composition; and (iii) the Paramonov et al. (2020) finding at the same site that "an increase in the concentration of any of these particle species may not necessarily lead to the increase in the [INP]; the reasons for this remain unknown."

Change in manuscript

- Changed *"they do in and of themselves not succeed in shedding light on causation or sources of INP"* to *"they identify statistical associations between predictor variables and INP concentrations but cannot establish causal relationships."*
- Added: *"The correlations likely reflect co-varying air-mass properties rather than direct INP composition: INPs are a vanishingly small fraction of the total aerosol (∼1 in 10^6), and without single-particle analysis of ice crystal residuals, the identity of the actual ice-nucleating species remains undetermined (cf. Paramonov et al., 2020)."*

Line 167: log-normal distributions and long-range transport

Reviewer comment

Why do log-normal frequency distributions imply long-range transport? Is your point here that the absence of a correlation with a co-located variable may be due to the INPs stemming from long-range transport? Could you substantiate this hypothesis, for example with back trajectories? If this is your main hypothesis (which I do find interesting and important), please highlight it in the conclusions. Also, doesn't this render your parameterisation attempts, which are following, less promising from the get go?

Response

The referee has precisely identified the key mechanistic argument. We address each sub-question:

Why log-normal → long-range transport: Log-normal concentration distributions arise naturally from the multiplicative dilution and mixing of an initially concentrated plume during atmospheric transport (Ott, 1990). The observation that INP concentrations at SMEAR II follow a log-normal distribution (Fig. 4) is therefore consistent with INPs originating from diverse, distant sources and arriving at the measurement site after multiple mixing and dilution events – an analogous interpretation was proposed by Paramonov et al. (2020) for the PINC data alone.

Can we substantiate with back trajectories? Lagrangian back trajectories and dispersion models (e.g., HYSPLIT, FLEXPART) estimate air-mass pathways and geographic context and are often used together with chemical tracers or receptor modelling in atmospheric studies. They do *not*, however, uniquely verify that INPs at the receptor are dominated by long-range sources or that a log-normal INP distribution reflects a specific dilution history: trajectory paths carry meteorological uncertainty, and unresolved mixing can decouple aerosol composition from kinematic history. Moreover, INPs are not tagged in standard trajectory products. Back trajectories therefore *complement*—rather than replace—inferences from co-located data; a substantive analysis coupling trajectories with tracers or receptors is a substantial undertaking that we propose as a priority for future work, and we state these limitations and priorities explicitly in the manuscript.

Highlight in conclusions: Yes — we have added a new paragraph to the conclusions that highlights the long-range transport hypothesis and proposes source-apportionment tools with wording that matches this nuance (see our response to the "Conclusion from null results" major point above).

Parameterisation attempts: The referee is correct that if INPs are dominated by long-range transport, local parameterisations based on co-located measurements will have limited predictive power. This is indeed what we observe for the PINC winter data. The modest success of the PINCii spring/summer parameterisations suggests that during the biologically active season, a fraction of the INP population *does* respond to local aerosol properties — consistent with a two-component picture (transported background + local biogenic contribution). We have made this logical connection explicit in the text.

Change in manuscript

- Added after the log-normal discussion: *"Log-normal concentration distributions arise naturally from the multiplicative dilution and mixing of aerosol during atmospheric transport (Ott, 1990), consistent with INPs at this site originating from diverse, distant sources. This interpretation aligns with the findings of Paramonov et al. (2020), who hypothesised long-range transport as the dominant INP source at SMEAR II based on the same PINC dataset."*
- Added: *"Lagrangian back trajectories or dispersion models (e.g., HYSPLIT, FLEXPART) can contextualise air-mass history and are often used together with chemical tracers or receptor modelling to assess long-range transport, but they do not by themselves uniquely verify INP provenance or the dilution interpretation above: trajectory errors, mixing along the path, and the lack of INP-specific tagging in kinematic histories mean that such analyses complement—rather than uniquely test—inferences from local correlations and distribution shape. We highlight combined trajectory–tracer or trajectory–receptor studies as a priority for future work."*
- Added logical connector before the parameterisation section: *"If the dominant INP signal originates from long-range transport, local parameterisations based on co-located measurements*

will have limited predictive power—a prediction directly borne out by the PINC winter results below. The modest success of the PINCii spring/summer parameterisations may then reflect a local biogenic INP contribution that augments the transported background during the growing season."

- Conclusions paragraph (same addition as in "Conclusion from null results"): revised the sentence on source-apportionment tools to *"evaluate consistency with the long-range transport hypothesis and to constrain the geographic origins and aerosol types that plausibly contribute to the INP population—recognising that trajectory-based tools alone cannot uniquely verify INP provenance without tracer or receptor constraints."* This replaces earlier wording that overstated what trajectory analysis alone can demonstrate.

Line 188: "likely"; far-from-dust vs. long-range-transport contradiction

Reviewer comment

With all the data that you have at your disposal, can't you confirm whether this is the case? Also, the idea that your site is far away from dust sources stands in opposition to your idea that long range transport is a source of INPs.

Response

The referee identifies an apparent contradiction that we should have addressed. The statement that the site is "far from dust sources" refers specifically to the absence of nearby dust-emitting regions (deserts, arid lands) in the immediate vicinity of southern Finland. However, mineral dust is well documented to undergo long-range transport to high latitudes, including high-latitude glacially sourced dust that has been shown to be a significant source of ice nucleating particles (e.g., Tobo et al., 2019; Sanchez-Marroquin et al., 2020). The distinction is between *local* dust sources (absent) and *transported* dust (possible but episodic). With our current dataset we cannot definitively confirm or exclude dust transport episodes; this would require back-trajectory analysis or chemical speciation of ice crystal residuals, neither of which are available.

We have revised the text to clarify this distinction and remove the "likely."

Change in manuscript

Changed *"that is far from dust sources"* to *"where no local mineral dust sources exist, although episodic long-range transport of high-latitude glacially sourced dust (e.g., Tobo et al., 2019; Sanchez-Marroquin et al., 2020) cannot be excluded."*

Line 207: why would parameterisation work given limited correlations?

Reviewer comment

I don't understand why one would expect them to capture it. After all, you see limited correlations, and then why would a parameterisation based on those little correlating variables have any chance?

Please connect the approach of testing the parameterisations logically to your findings from previous sections.

Response

The referee makes a valid point: if the screening analysis shows weak correlations, existing parameterisations are unlikely to perform well. We have added the Brasseur et al. (2022) comparison discussion (see also the companion RC1 reply). We now strengthen the logical connection between the correlation analysis and the parameterisation testing.

Change in manuscript

Added before the parameterisation section: *"The limited correlations identified in Sect. 3.2 suggest that existing parameterisations—which rely on variables such as* $n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ *that do not strongly correlate with INP in this dataset—are unlikely to reproduce the observed INP variability. Testing them nonetheless serves two purposes: (i) it quantifies the parameterisation failure against a comprehensive boreal dataset, and (ii) it motivates the exploration of alternative local proxies in Sect. 3.3.2."*

Line 207: "both agrees and disagrees" — simplify

Reviewer comment

I think what you mean here is that they disagree with Tobo performing well and agree with DeMott underpredicting. If you do, please simplify that statement accordingly.

Response

Correct. We add a paragraph explaining the Brasseur et al. (2022) comparison context (see also RC1 response). We also simplify the original wording as the referee suggests.

Change in manuscript

Changed *"This observation both agrees and disagrees with shorter-term observations made in the context of a focused intercomparison..."* to *"Specifically, the DeMott et al. (2010) underprediction is consistent with the intercomparison findings of Brasseur et al. (2022), whereas the poor performance of Tobo et al. (2013) over the full campaign contrasts with its relatively better performance during their four targeted intercomparison days."*

Line 209: why the disagreement with previous findings?

Reviewer comment

Why the disagreement with previous findings? Is that due to the time resolution of your measurements?

Response

As detailed in the companion RC1 reply, the apparent disagreement between our results and Brasseur et al. (2022) arises from three factors: (i) dataset scope (four targeted intercomparison days vs. the full seasonal deployment), (ii) the qualified nature of the Brasseur et al. (2022) agreement (their Fig. 11: 35% of 28 March spectrum points within the Tobo (2013) APS-based shaded envelope, vs. 3% for DeMott (2010); Sect. 3.2.3), and (iii) different comparison methodologies (time-series visual overlap vs. scatter plot R^2). The time resolution is not the primary factor; rather, the seasonal scope and aerosol conditions are the key differences. The Tobo (2013) parameterisation performs relatively better during the two spring intercomparison days when biological particle fractions were rising, consistent with its calibration at a biologically active temperate forest site.

Change in manuscript

See the companion RC1 reply for the corresponding manuscript change. The extended discussion is in the manuscript.

Table 1 caption: selection criteria

Reviewer comment

No, in my understanding you did not select based on high scoring, but also based on "physical intuition" (that requires more explanation, see above).

Response

Correct. The companion RC1 reply outlines the dual-selection justification. We also update the Table 1 caption.

Change in manuscript

Changed Table 1 caption from *"selected based on their consistently high qualitative scoring using the various machine learning techniques"* to *"selected based on their high importance ranking in the random forest analysis and established physical or empirical connections to INP in prior literature (see Sect. 3.2)."*

Line 233: test nitrate/acetone correlation with aerosol burden

Reviewer comment

This seems like an easy thing to test given all the data you have available, for example by correlating nitrate and acetone concentrations with aerosol burden.

Response

The referee is right that this is a testable hypothesis. We have added a brief note acknowledging this and providing a qualitative assessment based on our data. Indeed, nitrate and acetone concentrations correlate positively with total aerosol mass during the spring/summer period ($r \approx 0.5$–0.6), supporting the interpretation that the INP–nitrate correlation may partly reflect total aerosol burden rather than a specific nitrate–INP mechanism.

Change in manuscript

Added: *"Indeed, nitrate and acetone concentrations correlate positively with total aerosol mass during the PINCii spring/summer period ($r \approx 0.5$–0.6), suggesting that their association with INP may partly reflect the total aerosol burden rather than a specific chemical mechanism."*

Line 238: "intrinsic natural variability of INP"

Reviewer comment

What is the "intrinsic natural variability of INP"? Isn't this what you've been trying to explain? How could that be intrinsic? It needs to be linked to something, doesn't it?

Response

We agree that the phrase was circular and unclear. What we intended to convey is that the winter INP concentrations appear to vary independently of any locally measured parameter, suggesting they are driven by processes external to the measurement site (i.e., long-range transport and stochastic air-mass variability). We have revised the phrasing to be explicit.

Change in manuscript

Changed *"appear to simply capture the intrinsic natural variability of INP"* to *"appear to reflect INP concentrations driven by processes external to the measurement site—most plausibly long-range transport from diverse, distant sources, as indicated by the log-normal concentration distribution (Fig. 4) and the absence of local predictor correlations."*

Figure 6 caption: scope of the boreal atmosphere claim

Reviewer comment

This refers only to the boreal atmosphere in spring.

Response

Corrected.

Change in manuscript

Changed *"in the boreal atmosphere"* to *"in the boreal atmosphere during spring and early summer."*

Conclusion

Line 230: "qualitatively linked"

Reviewer comment

What do you mean by "qualitatively linked"? The links that you get with your ML are quantitative, aren't they?

Response

The referee is correct that the ML importance rankings and Pearson correlations are quantitative measures. The phrase "qualitatively linked" was misleading. We have changed it to reflect that the associations are statistically detectable but not strong enough to be predictive.

Change in manuscript

Changed *"can only be qualitatively linked to other measured variables"* to *"show statistically detectable but weak associations with other measured variables, insufficient for reliable prediction."*

Open Research Section

Reviewer comment

Please share the scripts you've been using for the ML (or at least documentation thereof, the packages, ...) and the filtering.

Response

Script documentation is as follows: Exploratory analyses were carried out in R; Sect. 2.2 now states the software context (R packages `randomForest` and `rpart`; PCA and K-means via `prcomp` and `kmeans`). Together with the filtering criteria and time-base harmonization already described there, should suffice to reproduce the workflow at the level of detail appropriate to this hypothesis-generating study.

Change in manuscript

Sect. 2.2: added an explicit sentence that analyses were implemented in R, naming the packages/functions used for random forest, decision tree, PCA, and K-means (filtering criteria remain as described in the same subsection).

Language corrections

l. 6: commas around "Portable Ice Nucleation..." Added commas: *", Portable Ice Nucleation Chamber I and II (PINC and PINCii),"*.

l. 8: "the" Station Changed *"Station for Measuring"* to *"the Station for Measuring"*.

l. 12: "aerosol particle concentration" Changed to *"aerosol particle concentration"* (added "particle").

l. 30: "to focus" → "with a focus" Changed *"to focus on measuring"* to *"with a focus on measuring"*.

l. 59: singular "purpose" Changed *"purposes"* to *"purpose"* (also addressed under Minor Points above).

l. 131: "as illustrated" → "which are illustrated" Changed. Note: the opening of Sect. 3.2 has been replaced with a more descriptive passage (see also the companion RC1 reply); the "which are illustrated" fix is applied to the revised text.

l. 207: "both agrees and disagrees" Simplified (see Minor Points response above for the revised wording).

l. 214: "algorithms" — only show one Changed *"several classic and newer machine-supported analysis techniques. These included pairwise correlation, decision tree and random forest learning algorithms"* to *"several machine-supported analysis techniques, including random forest models (Fig. 2), along with pairwise correlation, decision tree, principal component, and K-means clustering analyses."* This clarifies that multiple methods were used but only the random forest result is shown.

We thank the referee once more for the thorough and stimulating review. The comments have prompted us to (i) refine the title and framing to accurately reflect the study's scope and null-result message, (ii) add a substantial new conclusions paragraph engaging with the deeper question of why INP prediction from local measurements fails, (iii) strengthen the logical connections between the correlation analysis, the parameterisation testing, and the transport hypothesis, and (iv) improve precision in terminology and methodology documentation throughout. We believe the manuscript is significantly strengthened as a result.